

# A Fast Algorithm for Sampling from the Posterior of a von Mises distribution\*

Peter G.M. Forbes<sup>a†</sup> and Kanti V. Mardia<sup>b,a</sup>

<sup>a</sup>Department of Statistics, University of Oxford,  
1 South Parks Road, Oxford OX1 3TG, UK

<sup>b</sup>Department of Statistics, School of Mathematics,  
University of Leeds, Leeds LS2 9JT, UK

June 24, 2014

## Abstract

Motivated by molecular biology, there has been an upsurge of research activities in directional statistics in general and its Bayesian aspect in particular. The central distribution for the circular case is von Mises distribution which has two parameters (mean and concentration) akin to the univariate normal distribution. However, there has been a challenge to sample efficiently from the posterior distribution of the concentration parameter. We describe a novel, highly efficient algorithm to sample from the posterior distribution and fill this long-standing gap.

## 1 Introduction

There has been renewed interest in directional Bayesian analysis in view of its fundamental applications to molecular biology [4, 8, 12]. Due to chemical constraints on the bonds of biomolecules, the geometry of these molecules can be described by a set of angles. Other applications include locating and tracking an electric signal [10] and the analysis of forensic fingerprint evidence [7]. All these applications involve circular data which is naturally modelled by the von Mises distribution.

The probability density function of the von Mises distribution with mean  $\mu \in S^1$  on the unit circle and concentration parameter  $\kappa \geq 0$  is given by [13]

$$p(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp \{ \kappa \cos(\theta - \mu) \},$$

---

\*This is an author-created, un-copyedited version of an article accepted for publication in the *Journal of Statistical Computation and Simulation*. The Version of Record is available online at <http://dx.doi.org/doi:10.1080/00949655.2014.928711>.

<sup>†</sup>Corresponding author. Email: [forbes@stats.ox.ac.uk](mailto:forbes@stats.ox.ac.uk)

where  $I_m(\cdot)$  is the modified Bessel function of the first kind and order  $m$ . The circular variance can be described by  $1 - r(\kappa)$  where  $r(\kappa) = I_1(\kappa)/I_0(\kappa)$ . Let  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$  be a vector of observations from a von Mises distribution. When a conjugate prior is used, the posterior distribution of the mean  $\mu \mid \boldsymbol{\theta}, \kappa$  is itself von Mises, and can be easily sampled via [3].

Let  $\pi(\kappa) \propto I_0(\kappa)^{-a} \exp(-b\kappa)$  be the conjugate prior for the concentration. The posterior is

$$p(\kappa) = \frac{A}{I_0(\kappa)^\eta} \exp(-\eta\beta_0\kappa) \quad (1)$$

where  $\eta > 0$  and  $\beta_0 \in (-1, \infty)$  are observed constants; in this case  $\eta = a + n$  and  $\beta_0 = b/(a + n) - n^{-1} \sum_{i=1}^n \cos(\theta_i - \mu)$ . For the case  $\eta = 1$  and  $\beta_0 > 1$  the normalization constant is  $A = \sqrt{\beta_0^2 - 1}$ , however in general the normalization constant is intractable [11]. In this paper we shall call (1) the *Bessel exponential distribution*.

Existing algorithms to sample (1) tend to generate from approximate distributions [10] or have a large overhead of sampled auxiliary variables [6]. We present a new, extremely fast algorithm to sample from the Bessel exponential distribution.

For large  $\kappa$ ,  $I_0(\kappa)$  is approximately  $\exp(\kappa)/\sqrt{2\pi\kappa}$  [1, eq. 9.7.1]. Plugging this approximation into (1) yields a gamma density with shape  $\eta/2 + 1$  and rate  $\eta(\beta_0 - 1)$ .

This insight motivates us to use a gamma-based acceptance-rejection sampler for  $\kappa$ . However, the above approximation for  $I_0(\kappa)$  breaks down for small  $\kappa$ , and thus great care is needed to ensure our rejection sampler is efficient for all values of  $\kappa$ . We derive the optimal gamma-based proposal distribution and show that the resulting sampler has an acceptance probability of at least 0.7 for all  $\eta$  and  $\beta_0$ . The minimum acceptance probability of  $\approx 0.7$  occurs when the distribution is concentrated around  $\kappa = 0$ .

The algorithm is described in Section 2 and derived in Section 3. Enhancements are considered in Section 4 and the algorithm's efficiency is explored in Section 5.

## 2 The algorithm

As discussed above, we can approximate the Bessel exponential distribution with a gamma distribution. However, because the ratio of these densities diverges as  $\kappa \rightarrow 0$ , we cannot directly use a gamma proposal for our rejection sampler. Instead we propose values  $\kappa = x - \varepsilon$  where  $\varepsilon > 0$  and  $x$  has a gamma distribution. This is an application of Marsaglia's exact approximation procedure, see [14] for more details.

Using a shifted gamma proposal with shape  $\eta\alpha + 1 \geq 1$  and scale  $\eta\beta > 0$  leads to the envelope function

$$q(\kappa; \alpha, \beta, \varepsilon) = M(\alpha, \beta, \varepsilon)(\kappa + \varepsilon)^{\eta\alpha} \exp(-\eta\beta\kappa); \quad \kappa \geq 0$$

for  $p(\kappa)$ , where the amplitude  $M(\alpha, \beta, \varepsilon)$  is chosen to ensure the ratio  $p/q$  is bounded below one. We can generate a sample from the Bessel exponential distribution by generating a sample  $\kappa$  from  $q$  and accepting it with probability

$$\frac{p(\kappa)}{q(\kappa; \alpha, \beta, \varepsilon)} = \frac{A}{M(\alpha, \beta, \varepsilon)} \exp\{\eta g(\kappa; \alpha, \beta, \varepsilon)\}$$

where

$$g(\kappa; \alpha, \beta, \varepsilon) = (\beta - \beta_0)\kappa - \alpha \log(\kappa + \varepsilon) - \log I_0(\kappa),$$

$$M(\alpha, \beta, \varepsilon) = A \exp\{\eta g(\kappa_0; \alpha, \beta, \varepsilon)\}$$

and  $\kappa_0 = \operatorname{argmax}_{\kappa \geq 0} g(\kappa; \alpha, \beta, \varepsilon)$ . The samples generated via this procedure have the correct Bessel exponential distribution for any choice of the proposal parameters  $\alpha, \beta, \varepsilon$ , though the values of these parameters will affect the algorithm's efficiency.

In Section 3 we show that the approximate optimal choices for the proposal parameters are

$$\beta = \begin{cases} \beta_0 + 1 & \text{if } \beta_0 \leq 1/(4\eta) - 2/(3\sqrt{\eta}) \\ \beta_0 + r(\kappa_0) + \frac{1 - r(\kappa_0)}{1 + 40\eta\{\beta_0 - 1/(4\eta) + 2/(3\sqrt{\eta})\}^2} & \text{otherwise,} \end{cases}$$

$$\varepsilon = \frac{\kappa_0 \mathcal{W}_0\{c_3 \exp(c_3)\}}{c_3 - \mathcal{W}_0\{c_3 \exp(c_3)\}}, \quad \alpha = \{\beta - \beta_0 - r(\kappa_0)\}(\kappa_0 + \varepsilon)$$

where the terms  $\kappa_0$  and  $c_3$  are

$$\kappa_0 = \frac{1 - 1/\eta + 1/(2\eta^2)}{\eta\beta_0 + \sqrt{2\eta + \eta^2\beta_0^2}} + \frac{1 + 3/(2\eta) - 1/(4\eta^3)}{(\eta + 1)\beta_0 + \sqrt{2\eta + 1 + \eta^2\beta_0^2}},$$

$$c_3 = -\frac{\beta - \beta_0 - \log\{I_0(\kappa_0)\}/\kappa_0}{\beta - \beta_0 - r(\kappa_0)},$$

and  $\mathcal{W}_0(\cdot)$  is the principal branch of the Lambert W function defined as  $t = \mathcal{W}_0(t) \exp\{\mathcal{W}_0(t)\}$  for  $\mathcal{W}_0(t) > -1$ .

The acceptance-rejection algorithm to generate a sample from the Bessel exponential distribution proceeds as follows:

1. Find efficient proposal parameters  $\alpha, \beta, \varepsilon$ , which will depend on  $\eta, \beta_0$ .
2. Draw  $x$  from a gamma distribution with shape  $\eta\alpha + 1$  and rate  $\eta\beta$ .
3. Draw  $u$  from a Uniform distribution on  $[0, 1]$ .
4. Accept  $\kappa = x - \varepsilon$  if  $\log u < \eta g(\kappa; \alpha, \beta, \varepsilon) - \eta g(\kappa_0; \alpha, \beta, \varepsilon)$ , else go to 2.

The detailed procedure is described in Algorithm 1. When implementing the algorithm, both the Bessel functions and the Lambert W function can be computed using software such as the General Scientific Library [9] or its R wrapper, the CRAN package **gsl**. In practice it is often possible to avoid computing these functions, as we show in Section 4.

### 3 Derivation of the algorithm

We will now derive the optimal parameters  $\alpha, \beta, \varepsilon$  for the proposal distribution of  $x = \kappa - \varepsilon$  which follows a gamma distribution with shape  $\eta\alpha + 1$  and rate  $\eta\beta$ . We do so by maximizing the expected probability of acceptance

$$E_{x|\alpha, \beta}[I(x \geq \varepsilon) \exp\{\eta g(x - \varepsilon; \alpha, \beta, \varepsilon) - \eta g(\kappa_0; \alpha, \beta, \varepsilon)\}]$$

$$= \frac{(\eta\beta)^{\eta\alpha+1}}{\Gamma(\eta\alpha+1)} \exp\{-\eta\beta\varepsilon - \eta g(\kappa_0; \alpha, \beta, \varepsilon)\} \int_0^\infty \frac{\exp(-\eta\beta_0\kappa)}{I_0(\kappa)^\eta} d\kappa \quad (2)$$

---

**Algorithm 1** Rejection sampler for the Bessel exponential distribution

---

```

1:                                     ▷ Initialization: find parameters for proposal distribution
2:  $\kappa_L \leftarrow 2 / \left( \eta\beta_0 + \sqrt{2\eta + \eta^2\beta_0^2} \right)$ 
3:  $\kappa_U \leftarrow (2 + 1/\eta) / \left\{ (\eta + 1)\beta_0 + \sqrt{2\eta + 1 + \eta^2\beta_0^2} \right\}$ 
4:  $c_1 = 1/2 + \{1 - 1/(2\eta)\}/2\eta$ 
5:  $\kappa_0 \leftarrow (1 - c_1)\kappa_L + c_1\kappa_U$ 
6:  $i_0 \leftarrow I_0(\kappa_0)$ 
7:  $r \leftarrow I_1(\kappa_0)/i_0$ 
8:  $c_2 \leftarrow 1/(4\eta) - 2/(3\sqrt{\eta})$ 
9: if  $\beta_0 \leq c_2$  then
10:    $\beta \leftarrow \beta_0 + 1$ 
11: else
12:    $\beta \leftarrow \beta_0 + r + (1 - r)/\{1 + 40\eta(\beta_0 - c_2)^2\}$ 
13: end if
14:  $c_3 \leftarrow \{\log(i_0)/\kappa_0 - \beta + \beta_0\}/(\beta - \beta_0 - r)$ 
15:  $c_4 \leftarrow \mathcal{W}_0\{c_3 \exp(c_3)\}$                                      ▷ Lambert's W function, see Section 4
16:  $\varepsilon \leftarrow c_4\kappa_0/(c_3 - c_4)$ 
17:  $\alpha \leftarrow (\beta - \beta_0 - r)(\kappa_0 + \varepsilon)$ 
18:                                     ▷ Perform rejection sampling
19: repeat
20:    $x \leftarrow$  sample from a Gamma( $\eta\alpha + 1, \eta\beta$ ) left-truncated at  $\varepsilon$ 
21:    $\kappa \leftarrow x - \varepsilon$ 
22:    $u \leftarrow$  sample from a Uniform(0, 1)
23: until  $\log(u)/\eta < (\beta - \beta_0)(\kappa - \kappa_0) - \alpha \log\{(\kappa + \varepsilon)/(\kappa_0 + \varepsilon)\} - \log\{I_0(\kappa)/i_0\}$ 
24: return  $\kappa$ 

```

---

over  $\alpha, \beta, \varepsilon, \kappa_0$  subject to the constraint  $\kappa_0 = \operatorname{argmax}_{\kappa \geq 0} g(\kappa; \alpha, \beta, \varepsilon)$ . In order for the maximum to be finite as  $\kappa \rightarrow \infty$  we require  $\beta \leq \beta_0 + 1$ .

By taking logs we see that maximizing (2) with respect to  $\kappa$  is equivalent to maximizing

$$h(\kappa_0; \alpha, \beta, \varepsilon) = (\alpha + \eta^{-1}) \log(\eta\beta) - \eta^{-1} \log \Gamma(\eta\alpha + 1) - \beta\varepsilon - g(\kappa_0; \alpha, \beta, \varepsilon). \quad (3)$$

The constraint  $\kappa_0 = \operatorname{argmax}_{\kappa \geq 0} g(\kappa; \alpha, \beta, \varepsilon)$  implies either  $\kappa_0 = 0$  or  $\frac{d}{d\kappa} g(\kappa = \kappa_0; \alpha, \beta, \varepsilon) = 0$ . The Lagrangians for constrained optimization corresponding to these conditions are  $h + \lambda\kappa_0$  and  $h + \lambda \frac{d}{d\kappa} g$ , neither of which have interior critical points over  $(\alpha, \beta, \varepsilon, \lambda)$  because the  $\alpha$  and  $\varepsilon$  derivatives have no common root. Thus the optimal parameters must lie on the boundary of the parameter space. An examination of the boundaries show that the maximum satisfies  $\frac{d}{d\kappa} g(\kappa = \kappa_0; \alpha, \beta, \varepsilon) = 0$  and  $g(\kappa_0; \alpha, \beta, \varepsilon) = g(0; \alpha, \beta, \varepsilon)$ . Intuitively, this says that for any  $\kappa_0$  we should pick  $\varepsilon$  as small as possible while still having  $\kappa_0$  be the maximizer of  $g$ . Thus our Lagrangian is

$$L(\kappa_0; \alpha, \beta, \varepsilon, \lambda_1, \lambda_2) = h(\kappa_0; \alpha, \beta, \varepsilon) + \lambda_1 \frac{dg(\kappa = \kappa_0; \alpha, \beta, \varepsilon)}{d\kappa} + \lambda_2 \{g(\kappa_0; \alpha, \beta, \varepsilon) - g(0; \alpha, \beta, \varepsilon)\}.$$

Our optimal parameters are either a critical point of  $L$  or lie on one or more of the boundaries  $\alpha = 0, \varepsilon = 0$  or  $\beta = \beta_0 + 1$ . If either  $\alpha = 0$  or  $\varepsilon = 0$  then direct

differentiation shows the maximum occurs when  $\alpha = 0, \varepsilon = 0, \beta = \beta_0 + r(\kappa)$  and  $\kappa_0$  is the unique positive root of  $\beta - 1/(\eta\kappa)$ . We shall see this is a limiting case of the critical point solution. The only other boundary is  $\beta = \beta_0 + 1$ ; we shall see this is the solution when  $\beta_0$  is close to  $-1$ .

To find the critical points of  $L$ , we start by setting the derivatives with respect to  $\alpha, \varepsilon, \lambda_1$  and  $\lambda_2$  to zero and rearranging yields the optimal parameters as functions of  $\kappa_0$  and  $\beta$ . This yields

$$\begin{aligned}\alpha &= \{\beta - \beta_0 - r(\kappa_0)\}(\kappa_0 + \varepsilon), \\ \varepsilon &= \frac{\kappa_0 \mathcal{W}_0\{c_3 \exp(c_3)\}}{c_3 - \mathcal{W}_0\{c_3 \exp(c_3)\}}, \quad c_3 = -\frac{\beta - \beta_0 - \log\{I_0(\kappa_0)\}/\kappa_0}{\beta - \beta_0 - r(\kappa_0)}, \\ \lambda_2 &= \frac{\Psi(\eta\alpha + 1) - \log\{\eta\beta(\kappa_0 + \varepsilon)\} - 1 + \beta(\kappa_0 + \varepsilon)/\alpha}{\log(1 + \kappa_0/\varepsilon) - \kappa_0/\varepsilon}, \\ \lambda_1 &= \frac{\kappa_0 + \varepsilon}{\alpha\varepsilon} \{\beta\varepsilon^2 + (\kappa_0\beta - \alpha)\varepsilon + \alpha\lambda_2\kappa_0\},\end{aligned}\tag{4}$$

where  $\Psi(x) = \frac{d}{dx} \log \Gamma(x)$  is the digamma function. We must have  $\beta > \beta_0 + r(\kappa_0)$  so that  $\alpha$  and  $\varepsilon$  are positive. Notice that the limit  $\beta \rightarrow \beta_0 + r(\kappa_0)$  corresponds to the boundary case  $\alpha = \varepsilon = 0$  discussed above.

The above equations give all optimal parameters in terms of  $\beta$  and  $\kappa_0$ . Note that since constraint  $\frac{d}{d\kappa} g(\kappa; \alpha, \beta, \varepsilon) = 0$  at  $\kappa = \kappa_0$  is satisfied whenever  $\alpha = \{\beta - \beta_0 - r(\kappa_0)\}(\kappa_0 + \varepsilon)$ , we are free to choose alternative, sub-optimal values for the other parameters if the true optimal values are too difficult to compute. We shall explore this in Section 4 when we use an approximation for the Lambert W function.

Finally, we find the optimal  $\beta$  as follows. This value must either lie on the boundary  $\beta = \beta_0 + 1$  or else satisfy

$$\frac{\partial L(\kappa_0; \alpha, \beta, \varepsilon, \lambda_1, \lambda_2)}{\partial \beta} = (\alpha + 1/\eta)/\beta - (\kappa_0 + \varepsilon) + \lambda_1 - \lambda_2\kappa_0 = 0.\tag{5}$$

Unfortunately plugging (4) into (5) and solving for  $\beta$  as a function of  $\kappa_0$  alone is analytically intractable. However, one can check that  $\partial L/\partial \beta$  decreases from positive infinity at  $\beta = \beta_0 + r(\kappa_0)$  to negative infinity as  $\beta \rightarrow \infty$ . Since all admissible  $\beta$  lie in the finite interval

$$\max\{0, \beta_0 + r(\kappa_0)\} < \beta < \beta_0 + 1,$$

we can easily find the optimal  $\beta$  through any standard one-dimensional root-finding algorithm. If the root lies to the right of  $\beta_0 + 1$ , the optimal value is  $\beta = \beta_0 + 1$ .

We plug the optimal  $\beta$  into (4) to find all of our optimal parameters in terms of  $\kappa_0$ . Doing this for each  $\kappa_0$  and plugging the resulting parameters  $(\alpha, \beta, \varepsilon)$  into  $h(\kappa_0; \alpha, \beta, \varepsilon)$  yields a function which we numerically maximize over  $\kappa_0$ . Let  $\kappa^*$  be the optimal value of  $\kappa_0$  and let  $(\alpha^*, \beta^*, \varepsilon^*)$  be the optimal parameters corresponding to  $\kappa^*$ . These are the desired parameters that maximize the expected acceptance probability (2).

The above numeric maximizations for  $\beta$  and  $\kappa_0$  may be acceptable when  $\eta$  and  $\beta_0$  are known a priori. However, they are computationally prohibitive in the standard Monte Carlo case where we wish to generate many samples from the Bessel exponential distribution with different values of  $\eta$  and  $\beta_0$  for each

sample. Thus our next task is to approximate the optimal parameters with easily computable functions of  $\eta$  and  $\beta_0$ .

For all  $\eta$  and  $\beta_0$ ,  $\kappa^*$  is well approximated by  $\kappa_a$ , the positive root of  $\beta_0 + r(\kappa) - 1/(\eta\kappa)$ ; indeed  $\kappa_a$  is the exact optimum in the boundary case  $\alpha = \varepsilon = 0$ . To approximate  $\kappa_a$  we use the bounds [2, eq. 11],

$$\frac{\kappa}{1 + \sqrt{1 + \kappa^2}} \leq r(\kappa) \leq \frac{\kappa}{\sqrt{4 + \kappa^2}}.$$

Rearranging these bounds shows that  $\kappa_L \leq \kappa_a \leq \kappa_U$  where

$$\kappa_L = 2 \left( \eta\beta_0 + \sqrt{2\eta + \eta^2\beta_0^2} \right)^{-1}, \quad \kappa_U = (2+1/\eta) \left\{ (\eta+1)\beta_0 + \sqrt{2\eta + 1 + \eta^2\beta_0^2} \right\}^{-1}.$$

These bounds are relatively tight, we found that the convex combination  $(1 - c_1)\kappa_L + c_1\kappa_U$  with  $c_1 = 1/2 + \{1 - 1/(2\eta)\}/2\eta$  provides a good approximation to  $\kappa_a$  and hence to  $\kappa^*$ .

The parameter  $\beta^*$  is exactly equal to  $\beta_0 + 1$  when  $\beta_0$  is close to its lower bound of negative one. For  $\beta_0$  sufficiently large,  $\beta^*$  drops from its upper limit  $\beta_0 + 1$  towards its lower limit  $\beta_0 + r(\kappa^*)$ . The transition between the two limits is very rapid for  $\eta > 10$ . We achieve good accuracy with the approximation

$$\beta^* \approx \begin{cases} \beta_0 + 1 & \text{if } \beta_0 < c_2 \\ \beta_0 + r(\kappa^*) + \frac{1 - r(\kappa^*)}{1 + 40\eta(\beta_0 - c_2)^2} & \text{otherwise,} \end{cases}$$

where  $c_2 = 1/(4\eta) - 2/(3\sqrt{\eta})$ . This approximation is very good when  $\eta$  is large or when  $|\beta_0| > 0.1$ . A slower, more precise approximation for  $\beta^*$  may lead to parameters which provide better efficiency for small  $\eta$  and  $\beta_0 \approx 0$ ; we address this in Section 5.

Given these approximations of  $\kappa^*$  and  $\beta^*$ , the parameters  $\alpha^*$  and  $\varepsilon^*$  are given by (4).

## 4 Further speed enhancements

The truncated gamma on line 20 can be sampled using Dagpunar's algorithm [5]. Alternatively, one can use a standard gamma sampling algorithm such as Marsaglia–Tsang [15] and reject when  $x < \varepsilon$ . Indeed, the Marsaglia–Tsang algorithm is itself a rejection sampler with a Gaussian proposal, and its rejection step can be combined with the rejection step on line 23 for an additional speed-up.

The function  $\mathcal{W}_0$  on line 15 can be approximated by (Winitzki, [16])

$$\mathcal{W}_0(t) = \frac{et}{1 + \{(2et + 2)^{-1/2} + (e - 1)^{-1} - 2^{-1/2}\}^{-1}} \quad (6)$$

with no noticeable drop in the expected probability of acceptance.

Finally, we can implement the simple squeezes

$$\begin{aligned} I_0(\kappa) &< \{1 + 1/(2\kappa)\} \exp(\kappa)/\sqrt{2\pi\kappa} : & \kappa > 0 \\ I_0(\kappa) &> \exp(\kappa)/\sqrt{2\pi\kappa} : & \kappa > 0.259, \end{aligned}$$

to avoid computing the Bessel function within the rejection loop on line 23. Specifically, the loop on lines 19 to 24 can be replaced with Algorithm 2.

---

**Algorithm 2** Optimized loop replacing lines 19–24 of algorithm 1

---

```
 $c_5 = \log(i_0)$ 
loop
   $x \leftarrow$  sample from a  $\text{Gamma}(\eta\alpha + 1, \eta\beta)$  left-truncated at  $\varepsilon$ 
   $\kappa \leftarrow x - \varepsilon$ 
   $c_6 \leftarrow \frac{1}{2} \log(2\pi\kappa) - \kappa$ 
   $u \leftarrow$  sample from a  $\text{Uniform}(0, 1)$ 
   $v \leftarrow \log(u)/\eta - (\beta - \beta_0)(\kappa - \kappa_0) + \alpha \log\{(\kappa + \varepsilon)/(\kappa_0 + \varepsilon)\} - c_5$ 
  if  $\kappa < 0.258$  or  $v < c_6$  then
    if  $v < c_6 - \log\{1 + 1/(2\kappa)\}$  or  $v < -\log\{I_0(\kappa)\}$  then
      return  $\kappa$ 
end loop
```

---

## 5 Efficiency analysis

We now analyze the efficiency of Algorithm 1. When using the Winitzki approximation, the initial setup (lines 2–17) involve arithmetic operations, four square roots, two Bessel function evaluations, one logarithm and one exponentiation. In total this setup requires approximately 70 microseconds on a 2.4GHz Intel i5 computer when using the R package **gsl**. Implementation in a lower-level language would increase the speed significantly.

Each iteration of the rejection loop requires a gamma sample, a uniform sample, between three and five logarithms, and in the worst case a Bessel function evaluation. The squeeze in Algorithm 2 does a good job of avoiding the Bessel computation most of the time, and each iteration of the loop requires approximately 10 microseconds. Most of these iterations are accepted, and the algorithm, implemented in R, yields approximately 80,000 von Mises samples per second when  $\eta = 10$  and  $\beta_0$  is drawn uniformly over  $(-1, 1)$ . When using a compiled language such as C++, the algorithm yields over one million samples per second.

In Figure 1 we plot the expected probabilities of acceptance as functions of  $\eta$  and  $\beta_0$ . The figures were generated by numerically integrating the expected probability of acceptance (2) for each  $\eta = 1, 5, 10, 100$  and for each of 2000 equally spaced values of  $\beta_0 \in (-1, 1)$ .

There is a noticeable dip in efficiency near  $\beta_0 = 0$ . Recalling that  $\beta_0 = -n^{-1} \sum_{i=1}^n \cos(\theta_i - \mu)$ , we see that this region corresponds to diffused  $\theta_i$ , i.e. the true  $\kappa$  is near zero. This is precisely the region where our Bessel function approximation fails, so this drop is to be expected. Fortunately the drop in efficiency is not severe and our efficiency remains above 0.7 for all  $\eta$  and  $\beta_0$ .

From Figure 1 we see that our algorithm with the approximate optimal parameters does noticeably worse than the numerically computed true optimal parameters when  $\beta_0 \approx 0$ . This corresponds to the transition region where the optimal  $\beta$  rapidly drops from its upper limit of  $\beta_0 + 1$  to its lower limit of  $\beta_0 + r(\kappa_0)$ . Our approximation of the optimal  $\beta$  is inaccurate in this transition region. A more sophisticated approximation of the optimal  $\beta$  would increase the algorithm's efficiency, however, the region  $\kappa \approx 0$  is not usually an area of primary interest and we prefer to use the faster approximation.

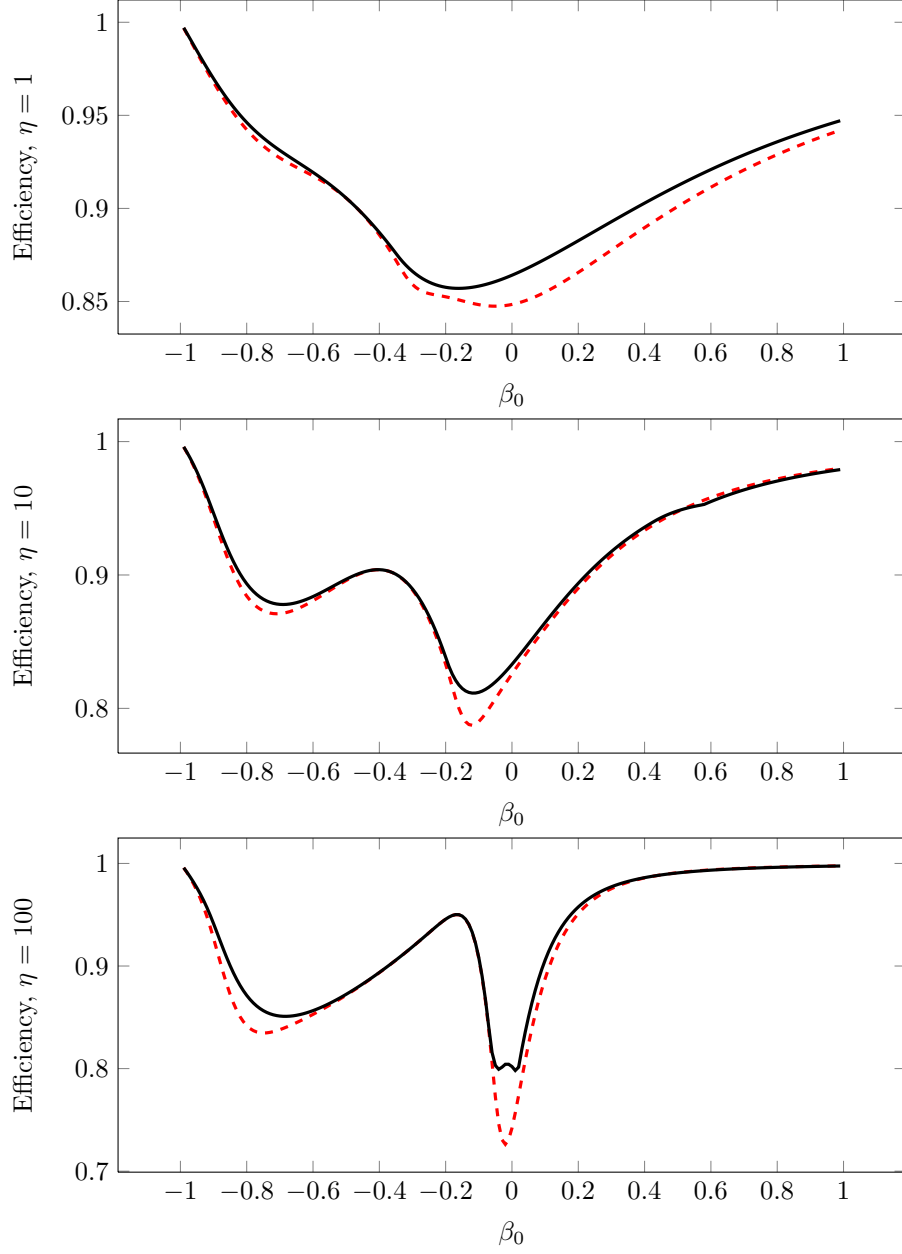


Figure 1: Expected probability of accepting a proposed  $\kappa$  (eq. 2), for  $\eta = 1, 10, 100$ . Red dashed lines correspond to algorithm 1 with the Winitzki approximation, black lines correspond to the numerically computed true optimal values for  $\beta$  and  $\kappa_0$ .



## 6 Conclusions

We have described a highly efficient algorithm to sample from the Bessel exponential distribution. It is suitable for any application where one wishes generate samples from the posterior distribution for the concentration parameter of the von Mises distribution.

## References

- [1] Abramowitz, M. and I. Stegun (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Government Printing Office, Washington D.C.
- [2] Amos, D. (1974). Computation of modified Bessel functions and their ratios. *Mathematics of Computation* 28(125).
- [3] Best, D. J. and N. I. Fisher (1979). Efficient simulation of the von Mises distribution. *Journal of the Royal Statistical Society Series C* 28(2), 152–157.
- [4] Boomsma, W., K. V. Mardia, C. C. Taylor, J. Ferkinghoff-Borg, A. Krogh, and T. Hamelryck (2008). A generative, probabilistic model of local protein structure. *Proceedings of the National Academy of Sciences* 105, 8932–8937.
- [5] Dagpunar, J. (1978). Sampling of variates from a truncated gamma distribution. *Journal of Statistical Computation and Simulation* 8, 59–64.
- [6] Damien, P. and S. Walker (1999). A full Bayesian analysis of circular data using the von Mises distribution. *The Canadian Journal of Statistics* 27(2), 291–298.
- [7] Forbes, P. G. M. and S. Lauritzen (2013). Fingerprint analysis using Bayesian alignment. In *Proceedings of the Leeds Annual Statistics Research Workshop*. Available: <http://www1.maths.leeds.ac.uk/statistics/workshop/lasr2013/proceedings/Forbes.pdf>.
- [8] Frellsen, J., I. Moltke, M. Thiim, K. V. Mardia, J. Ferkinghoff-Borg, and T. Hamelryck (2009). A probabilistic model of local RNA 3-D structure. *Public Library of Science Computational Biology* 5, 1–11.
- [9] Galassi, M., J. Davies, J. Theiler, B. Gough, and G. Jungman (2009). *GNU Scientific Library Reference Manual, Third Edition*. Network Theory Ltd.
- [10] Guttorp, P. and R. A. Lockhart (1988). Finding the location of a signal: A Bayesian analysis. *Journal of the American Statistical Association* 83(402), 322–330.
- [11] Mardia, K. V. (2007). On some recent advancements in applied shape analysis and directional statistics. In *Proceedings of the Leeds Annual Statistics Research Workshop*. Available: <https://www1.maths.leeds.ac.uk/statistics/workshop/lasr2007/proceedings/mardia.pdf>.
- [12] Mardia, K. V. (2013). Statistical approaches to three key challenges in protein structural bioinformatics. *Journal of the Royal Statistical Society Series C* 62, 487–514.
- [13] Mardia, K. V. and P. E. Jupp (1999). *Directional Statistics* (2nd ed.). Chichester, UK: Wiley.
- [14] Marsaglia, G. (1984). The exact-approximation method for generating random variables in a computer. *Journal of the American Statistical Association* 79(385), 218–221.
- [15] Marsaglia, G. and W. W. Tsang (2000). A simple method for generating gamma variables. *ACM Transactions on Mathematical Software* 26(3), 363–372.
- [16] Winitzki, S. (2003). Uniform approximations for transcendental functions. In *Computational Science and Its Applications*, Volume 2667 of *Lecture Notes in Computer Science*, pp. 780–789. Springer.